30

Lessons in biostatistics

Norman E. Breslow
Department of Biostatistics
University of Washington, Seattle, WA

Today's medical journals are full of factual errors and false conclusions arising from lack of statistical common sense. Reflecting on personal experiences, I argue that statisticians can substantially improve medical science by informed application of standard statistical principles. Two specific areas are identified where lack of such input regularly produces faulty research. Statisticians are needed more than ever to bring rigor to clinical research.

30.1 Introduction

Biostatisticians develop and apply statistical concepts and methods to clinical medicine, to laboratory medicine and to population medicine or public health. During the fifty years since COPSS was organized, their work has become increasingly important. Major medical journals often insist on biostatistical review of submitted articles. Biostatistics graduates are in high demand for work in industry, government and academia. They occupy prominent positions as heads of corporations and universities, deans of schools of public health and directors of major research programs.

In spite of the heightened visibility of the profession, much of today's medical research is conducted without adequate biostatistical input. The result is not infrequently a waste of public resources, the promulgation of false conclusions and the exposure of patients to possible mistreatment. I describe a few of the more common episodes of flawed research with which I have come in contact, which involve "immortal time" in follow-up studies and lack of proper validation of discriminant rules. I discuss the lessons learned both from these episodes and more generally from my decades of work in childhood cancer. The primary focus of the chapter is on biostatistics in clinical medicine. Other chapters in this volume discuss the role of statistics in laboratory medicine, especially genetics, and in public health.

30.2 It's the science that counts

My introduction to biostatistics was in graduate school. During the school year a small group from the Stanford statistics department made the trek to the medical school for a weekly seminar. There we learned from medical faculty and our professors about the research problems on which they were collaborating. During the summer we took jobs with local research organizations. At weekly meetings back on campus, we presented the problems stemming from our work and got advice from each other and the professors on how to approach them.

One summer I worked at the state health department. There was considerable interest in the possibility of an infectious origin for leukemia and speculation that transmission of the putative infectious agent might occur between animals and humans. The health department was conducting a census of cancer occurrence in dogs and cats in Alameda county, and the epidemiologists wanted to evaluate possible space-time clustering of leukemia cases in people and in cats. The maps at their disposal, however, were inaccurate. Ascertainment of the geographic coordinates needed for quantitative analysis was subject to substantial error. My assignment was to read up on spatial statistical distributions and develop a measurement error model. I was having considerable difficulty.

I will never forget the stern advice I received from Professor Lincoln Moses following my presentation at the weekly meeting back on campus. "What you need is a good set of maps," he said. "Try the water company!" Obviously, in his mind, as later in mine, the best method of dealing with measurement error was to avoid it! Bradford Hill gave similar advice:

> "One must go and seek more facts, paying less attention to the techniques of handling the data and far more to the development and perfection of the methods of obtaining them." (Hill, 1953)

As it turned out, the East Bay Municipal Utilities District (EBMUD) had just completed a very extensive and costly mapping program. The maps were so accurate that you had to decide where in the residence to plot the case to determine the coordinates. Executives in charge of the program were delighted to learn that their maps would serve not only corporate interests but also those of public health. Instead of working on a statistical methods problem, I spent my remaining time that summer on administrative issues related to the use of the maps by the health department. A photograph of me with health department and EBMUD officials poring over the maps was published in the corporate magazine and hangs in my office today. The lesson learned was invaluable.

I had a similar experience shortly after my arrival at the University of Washington in 1968. Having applied for a position in the Mathematics Department, not realizing it was in the process of downsizing and discharging

most of its statisticians, I wound up as a biostatistician in the Medical School. My support came mainly from service as statistician to the Children's Cancer Study Group. In those days the MDs who chaired the protocol study committees sometimes compiled the data themselves (one dedicated researcher meticulously arranged the flow sheets on her living room floor) and sent me simple data summaries with a request for calculation of some standard statistic. I was appalled by the routine exclusion from randomized treatment comparisons of patients who had "inadequate trials" of chemotherapy due to early discontinuation of the assigned treatment regimen or early death. It was clear that a more systematic approach to data collection and analysis was needed.

My colleague Dick Kronmal, fortunately, had just developed a computer system to store and summarize data from longitudinal studies that generated multiple records per patient (Kronmal et al., 1970). This system was perfect for the needs of the children's cancer group. It allowed me to quickly establish a Data and Statistical Center both for the group and for the National Wilms Tumor Study (NWTS), whose steering committee I joined as a founding member in 1969. (Wilms is an embryonal tumor of the kidney that occurs almost exclusively in children.) Once again the lesson learned was that "development and perfection of the methods of obtaining the data" were at least as important to the overall scientific enterprise as were the statistical methods I subsequently helped develop to "handle" right censored survival data. Having me, as statistician, take control of data collection and processing, while sharing responsibility for data quality with the clinicians, made it easier for me to then also exercise some degree of control over which patients were included in any given analysis.

My experience was not unusual. The role of biostatisticians in cooperative clinical research was rapidly evolving as the importance of their contributions became more widely appreciated. It soon became commonplace for them to occupy leadership positions within the cooperative group structure, for example, as heads of statistics departments or as directors of independently funded coordinating centers.

A steady diet of service to clinical trial groups, however, can with time become tedious. It also interferes with production of the first-authored papers needed for promotion in academia. One way to relieve the tedium, and to generate the publications, is to get more involved in the science. For example, the biostatistician can propose and conduct ancillary studies that utilize the valuable data collected through the clinical trials mechanism. The first childhood leukemia study in which I was involved was not atypical in demonstrating that treatment outcomes varied much more with baseline host and disease characteristics, in this case age and the peripheral white blood cell count, than with the treatments the study was designed to assess (Miller et al., 1974). This result was apparently a revelation to the clinicians. They jumped on it to propose treatment stratification based on prognostic factor groups in subsequent trials, so that the most toxic and experimental treatments were reserved for those who actually needed them. Subsequently, I initiated several studies of

prognosis in Wilms tumor that resulted in greater appreciation for the adverse outcomes associated with tumor invasion of regional lymph nodes and ultimately to changes in the staging system. Fascinated by how well Knudson's (Knudson, Jr., 1971) 2-hit mutational model explained the genetic epidemiology of retinoblastoma, another embryonal tumor of a paired organ (in this case the eye rather than the kidney), I conducted studies of the epidemiology of Wilms tumor that provided strong evidence for genetic heterogeneity, an explanation for its lower incidence and younger ages-at-onset in Asians and a hypothesis regarding which survivors were at especially high risk of end stage renal disease in young adulthood (Breslow and Beckwith, 1982; Breslow et al., 2006; Lange et al., 2011). Since 1991, I have served as Principal Investigator on the NIH grant that funds the continued follow-up of NWTS survivors for "late effects" associated with Wilms tumor and its treatment. This study has occupied an increasingly important place in my research repertoire.

30.3 Immortal time

In my opening lecture to a class designed primarily for second year doctoral students in epidemiology, I state the First Rule of survival analysis: *Selection into the study cohort, or into subgroups to be compared in the analysis, must not depend on events that occur after the start of follow-up.* While this point may be obvious to a statistician, certainly one trained to use martingale arguments to justify inferences about how past history influences rates of future events, it was not obvious to many of the epidemiologists. The "immortal time" bias that results from failure to follow the rule has resulted, and continues regularly to result, in grossly fraudulent claims in papers published in the most prestigious medical journals.

My first exposure to the issue came soon after I started work with the children's cancer group. The group chair was puzzled by a recently published article that called into question the standard criteria for evaluation of treatment response in acute leukemia. These included the stipulation that patients with a high bone marrow lymphocyte count (BMLC) be excluded from the excellent response category. Indeed, a high BMLC often presaged relapse, defined as 5% or higher blast cells in the marrow. The article in question, however, reported that patients whose lymphocytes remained below the threshold level of 20% of marrow cells throughout the period of remission tended to have shorter remissions than patients whose BMLC exceeded 20% on at least one occasion. Although I knew little about survival analysis, and had not yet articulated the First Rule, I was familiar with random variation and the tendency of maxima to increase with the length of the series. Intuition suggested that the article's conclusion, that there was "no justification for excluding a patient

from complete remission status because of bone marrow lymphocytosis," was erroneous.

Accordingly, using new data from the children's cancer group, I attempted to convince my clinical colleagues that the reasoning was fallacious (Breslow and Zandstra, 1970). I first replicated the earlier findings by demonstrating that, when patients were classified into three categories according to the BMLC values observed during remission, the "remission duration" (progression-free survival) curve for the group with highest *maximum* BMLC was on top and that for the group with lowest maximum BMLC was on the bottom. When patients were classified according to the *average* of their BMLC values during remission, however, the ordering was reversed. Both comparisons were highly statistically significant. Of course, even the analysis based on average counts violated the First Rule. Nowadays one would employ time-dependent covariates or stratification to evaluate how the history of BMLC affected future relapse rates. The experience was a valuable lesson about the importance of "statistical thinking" in clinical research.

Many biostatisticians were sensitized to the issue of immortal time by Mitch Gail's critique of early claims of the efficacy of heart transplantation (Gail, 1972). To illustrate the problems with the statistical approach taken by cardiac surgeons in those days, he compared survival curves from time of admission as a transplant candidate according to whether or not the patient had subsequently received a transplant. He pointed out that patients who died early had less opportunity to receive a transplant, whereas those who did receive one were guaranteed, by definition, to have survived long enough for a suitable donor to be found. In effect, person-months of observation prior to transplant were unfairly subtracted from the total person-months for the control group, biasing their survival rate downwards, and added to the person-months for the transplant group, biasing their survival rate upwards. Correct accounting for the timing of transplant in the statistical comparison was subsequently undertaken by several statistical teams, for example, by use of time-dependent covariates in the Cox model (Crowley and Hu, 1977). When the data were properly analyzed, transplant as performed at the time was found to have little benefit.

Nick Day and I, in the section of our second IARC monograph (Breslow and Day, 1987) on allocation of person-time to time-dependent exposure categories, called attention to a fallacious claim of decreasing death rates with increasing duration of work in the polyvinyl-chloride industry. Here the investigators had contrasted standardized mortality ratios (of numbers of deaths observed to those expected from age-specific population rates) among workers employed for 0–14 versus 15+ years in the industry. Not only all deaths occurring beyond 15 years, however, but also all person-time accumulated by persons employed for 15+ years, had been allocated to the latter group. Day and I stated: "The correct assignment of each increment in person-years of follow-up is to that same exposure category to which a death would be assigned should it occur at that time." In other words, the first 15 years of employment

time for the vinyl-chloride workers whose employment continued beyond that point should have been assigned to the 0–14 group. When this correction was made, the 15+ year exposure group had a slightly higher mortality ratio than did the 0–14 year group.

Faculty at McGill University in Montréal, Canada, have repeatedly called attention to erroneous conclusions in the medical literature stemming from immortal time bias. One recent article takes issue with the finding that actors who won an Oscar lived on average nearly four years longer than those in a matched control group (Sylvestre et al., 2006). The authors pointed out that, as long ago as 1843, William Farr warned against the hazards of "classifying persons by their status at the end of follow-up and analyzing them as if they had been in these categories from the outset" (Farr, 1975). Farr continued

> "... certain professions, stations and ranks are only attained by persons advanced in years; and some occupations are followed only in youth; hence it requires no great amount of sagacity to perceive that 'the mean age at death' [$\cdots$] cannot be depended on in investigating the influence of occupation, rank and profession upon health and longevity."

Noting the relatively early ages at death of Cornets, Curates and Juvenile Barristers, he concluded wryly: "It would be almost necessary to make them Generals, Bishops and Judges — for the sake of their health."

Mistakes are made even when investigators are seemingly aware of the problem. A 2004 report in *The New England Journal of Medicine* examined the effect on survival of a delay in kidney transplantation among children with end stage renal disease. The authors stated:

> "Delay in kidney transplantation as a potential risk factor for early death was analyzed by comparing mortality among groups with different lengths of time until transplantation. To account for survival bias, delay as a predictor of death was analyzed beginning 2 years after the initiation of renal replacement therapy. There was no significant difference in mortality observed among groups with different lengths of time until transplantation (Fig 3)" (McDonald and Craig, 2004).

Close examination of their Figure 3, however, leads to a different conclusion. Survival curves from two years after onset of renal replacement therapy (dialysis or transplant) were shown separately for those with preemptive transplant (no delay), less than one-year delay and 1–2 years delay, categories based on information available at the start of follow-up at two years. They are in the anticipated order, with the survival outcomes best for those having had an immediate transplant followed in succession by those having had a 0–1 or 1–2 year delay. Had the authors simply added a fourth curve for those not yet transplanted by year 2, they would have found that it lay below the others. This would have confirmed the anticipated rank order in survival outcomes under the hypothesis that longer delay increased subsequent mortality. However, they mistakenly split the fourth group into those who never received a

transplant and those who did so at some point after two years. The survival curve for the "no transplant" group was far below all the others, with many deaths having occurred early on prior to a suitable donor becoming available, while the curve for the "$\geq$ 2 years" group was second from highest due to immortal time. The clear message in the data was lost. I have used this widely cited paper as the basis for several exam and homework questions. Students often find the lessons about immortal time to be the most important they learned from the class.

I mentioned earlier my dissatisfaction with the exclusion of patients with "inadequate trials" from randomized treatment comparisons, a policy that was widely followed by the children's cancer group when I joined it. Such "per protocol" analyses constitute another common violation of the First Rule. Exclusion of patients based on events that occur after the start of follow-up, in particular, the failure to receive protocol treatment, invites bias that is avoided by keeping all eligible patients in the study from the moment they are randomized. Analyses using all the eligible patients generate results that apply to a real population and that are readily compared with results from like studies. Attempts to clearly describe the fictitious populations to which the per protocol analyses apply are fraught with difficulty. My colleague Tom Fleming has thoughtfully discussed the fundamental principle that all patients be kept in the analysis following randomization, its rationale and its ramifications (Fleming, 2011).

30.4 Multiplicity

Whether from cowardice or good sense, I consciously strived throughout my career to avoid problems involving vast amounts of data collected on individual subjects. There seemed to be enough good clinical science to do with the limited number of treatment and prognostic variables we could afford to collect for the childhood cancer patients. The forays into the epidemiology of Wilms tumor similarly used limited amounts of information on gender, ages at onset, birth weights, histologic subtypes, precursor lesions, congenital malformations and the like. This allowed me to structure analyses using a small number of variables selected a priori to answer specific questions based on clearly stated hypotheses.

My successors do not have this luxury. Faced with the revolution in molecular biology, they must cope with increasingly high dimensional data in an attempt to assist clinicians deliver "personalized medicine" based on individual "omics" (genomics, epigenomics, proteomics, transcriptomics, metabolomics, etc.) profiles. I hope that widespread enthusiasm for the new technologies does not result in a tremendous expenditure of resources that does little to advance public health. This can be avoided if statisticians demand, and are given, a

meaningful role in the process. I am impressed by how eagerly my younger colleagues, as well as some of my peers, have responded to the challenge.

The problems of multiplicity were brought home to me in a forceful way when I read an article based on data from the 3rd and 4th NWTS trials supplied by our pathologist to a group of urologists and pathologists at the prestigious Brady Institute at Johns Hopkins Hospital (JHH); see Partin et al. (1994). Regrettably, they had not solicited my input. I was embarrassed that a publication based on NWTS data contained such blatant errors. For one thing, although our pathologist had supplied them with a case-control sample that was overweighted with children who had relapsed or had advanced disease at onset, they ignored the design and analysed the data as a simple random sample. Consequently their Kaplan–Meier estimates of progression-free survival were seriously in error, suggesting that nearly half the patients with "favorable histology" relapsed or died within five years of diagnosis, whereas the actual fraction who did so was about 11%.

A more grievous error, however, was using the same data both to construct and to validate a predictive model based on a new technology that produced moderately high dimensional quantitative data. Determined to improve on the subjectivity of the pathologist, the JHH team had developed a technique they called nuclear morphometry to quantify the malignancy grading of Wilms and other urologic tumors, including prostate. From the archived tumor slide submitted by our pathologist for each patient, they selected 150 blastemal nuclei for digitizing. The digitized images were then processed using a commercial software package known as Dyna CELL. This produced for each nucleus a series of 16 shape descriptors including, for example, area, perimeter, two measures of roundness and two of ellipticity. For each such measure 17 descriptive statistics were calculated from the distribution of 150 values: Mean, variance, skewness, kurtosis, means of five highest and five lowest values, etc. This yielded $16 \times 17 = 242$ nuclear morphometric observations per patient. Among these, the *skewness* of the nuclear roundness factor (SNRF) and the average of the lowest five values for ellipticity as measured by the feret diameter (distance between two tangents on opposite sides of a planar figure) method (LEFD) were found to best separate cases from controls, each yielding $p = .01$ by univariate logistic regression. SNRF, LEFD and age, a variable I had previously identified as an important prognostic factor, were confirmed by stepwise regression analysis as the best three of the available univariate predictors. They were combined into a discriminant function that, needless to say, did separate the cases from the controls used in its development, although only with moderate success.

TABLE 30.1
Regression coefficients (± SEs) in multivariate nuclear morphometric discriminant functions fitted to three data sets[†].

	Case-Control Sample		Prospective Sample
Risk Factor	NWTS + JHH ($n = 108$)[*]	NWTS Alone ($n = 95$)	NWTS ($n = 218$)
Age (yr)	.02	$.013 \pm .008$	$.017 \pm .005$
SNRF	1.17	$1.23 \pm .52$	$-.02 \pm .26$
LEFD	90.6	121.6 ± 48.4	$.05 \pm 47.5$

[†]From Breslow et al. (1999). Reproduced with permission.
[*]From Partin et al. (1994)

I was convinced that most of this apparent success was due to the failure to account for the multiplicity of comparisons inherent in the selection of the best 2 out of 242 measurements for the discriminant function. With good cooperation from JHH, I designed a prospective study to validate the ability of their nuclear morphometric score to predict relapse in Wilms tumor (Breslow et al., 1999). I identified 218 NWTS-4 patients who had not been included in the case-control study, each of whom had an archived slide showing a diagnosis by our pathologist of a Wilms tumor having the same "favorable" histologic subtype as considered earlier. The slides were sent to the JHH investigators, who had no knowledge of the treatment outcomes, and were processed in the same manner as for the earlier case-control study. We then contrasted results obtained by re-analysis of data for the 95 NWTS patients in the case-control study, excluding 13 patients from JHH who also had figured in the earlier report, with those obtained by analysis of data for the 218 patients in the prospective study.

The results, reproduced in Table 30.1, were instructive. Regression coefficients obtained using a Cox regression model fitted to data for the 95 NWTS patients in the original study are shown in the third column. They were comparable to those reported by the JHH group based on logistic regression analysis of data for the 95 NWTS plus 13 JHH patients. These latter coefficients, shown in the second column of the table, were used to construct the nuclear morphometric score. Results obtained using Cox regression fitted to the 218 patients in the prospective study, of whom 21 had relapsed and one had died of toxicity, were very different. As I had anticipated, the only variable that was statistically significant was the known prognostic factor age. Coefficients for the two nuclear morphometric variables were near zero. When the original nuclear morphometric score was applied to the prospective data, using the same cutoff value as in the original report, the sensitivity was reduced from

75% to 71% and the specificity from 69% to 56%. Only the inclusion of age in the score gave it predictive value when applied to the new data.

No further attempts to utilize nuclear morphometry to predict outcomes in patients with Wilms tumor have been reported. Neither the original paper from JHH nor my attempt to correct its conclusions have received more than a handful of citations. Somewhat more interest was generated by use of the same technique to grade the malignancy of prostate cancer, for which the JHH investigators identified the *variance* of the nuclear roundness factor as the variable most predictive of disease progression and disease related death. While their initial studies on prostate cancer suffered from the same failure to separate test and validation data that compromised the Wilms tumor case-control study, variance of the nuclear roundness factor did apparently predict adverse outcomes in a later prospective study.

Today the public is anxiously awaiting the anticipated payoff from their investment in omics research so that optimal medical treatments may be selected based on each patient's genomic or epigenomic make-up. Problems of multiplicity inherent in nuclear morphometry pale in comparison to those posed by development of personalized medicine based on omics data. A recent report from the Institute of Medicine (IOM) highlights the important role that statisticians and statistical thinking will play in this development (IOM, 2012). This was commissioned following the exposure of serious flaws in studies at Duke University that had proposed tests based on gene expression (microarray) profiles to identify cancer patients who were sensitive or resistant to specific chemotherapeutic agents (Baggerly and Coombes, 2009). Sloppy data management led to major data errors including off-by-one errors in gene lists and reversal of some of the sensitive/resistant labels. The corrupted data, coupled with inadequate information regarding details of computational procedures, made it impossible for other researchers to replicate the published findings. Questions also were raised regarding the integrity of the validation process. Ultimately, dozens of papers were retracted from major journals, three clinical trials were suspended and an investigation was launched into financial and intellectual/professional conflicts of interest.

The IOM report recommendations are designed to prevent a recurrence of this saga. They emphasize the need for evaluation of a completely "locked down" computational procedure using, preferably, an independent validation sample. Three options are proposed for determining when a fully validated test procedure is ready for clinical trials that use the test to direct patient management. To ensure that personalized treatment decisions based on omics tests truly do advance the practice of medicine, I hope eventually to see randomized clinical trials where test-based patient management is compared directly with current standard care.

30.5 Conclusion

The past 50 years have witnessed many important developments in statistical theory and methodology, a few of which are mentioned in other chapters of this COPSS anniversary volume. I have focussed on the place of statistics in clinical medicine. While this sometimes requires the creation of new statistical methods, more often it entails the application of standard statistical principles and techniques. Major contributions are made simply by exercising the rigorous thinking that comes from training in mathematics and statistics. Having statisticians take primary responsibility for data collection and management often improves the quality and integrity of the entire scientific enterprise.

The common sense notion that definition of comparison groups in survival analyses should be based on information available at the beginning of follow-up, rather than at its end, has been around for over 150 years. When dealing with high-dimensional biomarkers, testing of a well defined discriminant rule on a completely new set of subjects is obviously the best way to evaluate its predictive capacity. Related cross-validation concepts and methods have been known for decades. As patient profiles become more complex, and biology more quantitative, biostatisticians will have an increasingly important role to play in advancing modern medicine.

References

Baggerly, K.A. and Coombes, K.R. (2009). Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*, 3:1309–1334.

Breslow, N.E. and Beckwith, J.B. (1982). Epidemiological features of Wilms tumor — Results of the National Wilms Tumor Study. *Journal of the National Cancer Institute*, 68:429–436.

Breslow, N.E., Beckwith, J.B., Perlman, E.J., and Reeve, A.E. (2006). Age distributions, birth weights, nephrogenic rests, and heterogeneity in the pathogenesis of Wilms tumor. *Pediatric Blood Cancer*, 47:260–267.

Breslow, N.E. and Day, N.E. (1987). *Statistical Methods in Cancer Research II: The Design and Analysis of Cohort Studies*. IARC Scientific Publications. International Agency for Research on Cancer, Lyon, France.

Breslow, N.E., Partin, A.W., Lee, B.R., Guthrie, K.A., Beckwith, J.B., and Green, D.M. (1999). Nuclear morphometry and prognosis in favorable

histology Wilms tumor: A prospective reevaluation. *Journal of Clinical Oncology*, 17:2123–2126.

Breslow, N.E. and Zandstra, R. (1970). A note on the relationship between bone marrow lymphocytosis and remission duration in acute leukemia. *Blood*, 36:246–249.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, 72:27–36.

Farr, W. (1975). *Vital Statistics: A Memorial Volume of Selections from the Writings of William Farr*. Scarecrow Press, Metuchen, NJ.

Fleming, T.R. (2011). Addressing missing data in clinical trials. *Annals of Internal Medicine*, 154:113–117.

Gail, M.H. (1972). Does cardiac transplantation prolong life? A reassessment. *Annals of Internal Medicine*, 76:815–817.

Hill, A.B. (1953). Observation and experiment. *New England Journal of Medicine*, 248:995–1001.

Institute of Medicine (2012). *Evolution of Translational Omics: Lessons Learned and the Path Forward*. The National Acadamies Press, Washington, DC.

Knudson, A.G. Jr. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences*, 68:820–823.

Kronmal, R.A., Bender, L., and Mortense, J. (1970). A conversational statistical system for medical records. *Journal of the Royal Statistical Society, Series C*, 19:82–92.

Lange, J., Peterson, S.M., Takashima, J.R., Grigoriev, Y., Ritchey, M.L., Shamberger, R.C., Beckwith, J.B., Perlman, E., Green, D.M., and Breslow, N.E. (2011). Risk factors for end stage renal disease in non-WT1-syndromic Wilms tumor. *Journal of Urology*, 186:378–386.

McDonald, S.P. and Craig, J.C. (2004). Long-term survival of children with end-stage renal disease. *New England Journal of Medicine*, 350:2654–2662.

Miller, D.R., Sonley, M., Karon, M., Breslow, N.E., and Hammond, D. (1974). Additive therapy in maintenance of remission in acute lymphoblastic leukemia of childhood — Effect of initial leukocyte count. *Cancer*, 34:508–517.

Partin, A.W., Yoo, J.K., Crooks, D., Epstein, J.I., Beckwith, J.B., and Gearhart, J.P. (1994). Prediction of disease-free survival after therapy in Wilms tumor using nuclear morphometric techniques. *Journal of Pediatric Surgery*, 29:456–460.

Sylvestre, M.P., Huszti, E., and Hanley, J.A. (2006). Do Oscar winners live longer than less successful peers? A reanalysis of the evidence. *The Annals of Internal Medicine*, 145:361–363.